

Folding Kinetics of a Polymer

Štěpán Růžička, David Quigley, and Michael P. Allen

Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom

(Dated: September 30, 2011)

By simulating the first order globule-crystal transition of a flexible homopolymer chain, both by collision dynamics and Monte Carlo with non-kinetic moves, we show that the effective and the thermodynamic transition temperatures are different and we propose a way of quantifying the kinetic hindering. We then also observe that the top eigenvalue in the spectrum of the dynamical (contact or adjacency) matrix provides insight into the ensembles of folding and unfolding trajectories, and may be a suitable additional reaction coordinate for the folding transition of chain molecules.

PACS numbers: 64.70.km, 36.20.-r, 82.70.Dd, 02.70.Ns, 05.70.Ln

Keywords: POLYMERS, DYNAMICS

Dynamical simulations of phase transitions in simple chain molecules are important for understanding the underlying microscopic folding mechanisms of polymers and proteins. These simulations are usually computationally expensive, because the phases are often separated by a large free energy barrier, and the fluctuation leading to barrier crossing is a rare event on the molecular simulation timescale. Free energy barriers are usually studied by Monte Carlo (MC) simulation, and dynamical studies of barrier crossing are less common. Here, we use stochastic hard-sphere molecular dynamics, collision dynamics (CD) for short, to study the kinetics involved in the first-order crystallisation transition for a flexible homopolymer model, composed of bonded hard spheres with square-well nonbonded attractions, which has recently been intensively studied by MC [1–5]. By using CD, we avoid unphysical aspects of Monte Carlo (such as connectivity-altering moves improving sampling efficiency) and hence are able to observe the effects of kinetic bottlenecks in determining the rate, for a situation that mimics realistic dynamics.

In what follows, the polymer is represented by 128 hard spheres with two types of square-well potential for bonded and non-bonded pairs of monomers:

$$u_b(r_{i,i+1}) = \begin{cases} 0 & \sigma < r_{i,i+1} < \chi_b \sigma \\ +\infty & \text{otherwise,} \end{cases} \quad (1)$$

$$u_w(r_{ij}) = \begin{cases} +\infty & 0 < r_{ij} < \sigma \\ -\epsilon & \sigma < r_{ij} < \chi \sigma, \\ 0 & \chi \sigma < r_{ij} \end{cases} \quad |i - j| > 1, \quad (2)$$

where r_{ij} is the distance between centers of two monomers i and j , σ is the diameter of the bead, ϵ is the depth of the square well, and χ and χ_b are the relative widths of the square well and the nearest-neighbour bond respectively. In the MC studies [3] $\chi_b = 1$, but for simple hard sphere dynamics, we choose a slightly larger value $\chi_b = 1.04$; we have verified that this makes very little difference to the equilibrium properties. Taylor et al. [2, 3] have determined that an all-in-one ‘protein-like’ crystallisation from the expanded state occurs for

$\chi \lesssim 1.06$, and a two-step ‘polymer-like’ mechanism via a liquid-like globule, for $\chi \gtrsim 1.06$; we study the globule-crystal stage of the two-step process for χ values in this vicinity. All beads have equal mass m which we take equal to unity. Throughout, we work in reduced units: $\sigma = 1$, $\epsilon = 1$ and $k_B = 1$ (Boltzmann’s constant) so $T = k_B T / \epsilon$. Corresponding real values, for monomer beads corresponding to amino acids of the kind found in proteins, would be $m \approx 2 \times 10^{-25}$ kg, $\sigma \approx 6 \times 10^{-10}$ m, $\epsilon \approx 7 \times 10^{-22}$ J, and a time unit $\approx 10^{-11}$ s.

We have simulated the above chain model with both the WL and CD methods. The WL Monte Carlo move set [6, 7] consists of crankshaft, pivot, end-bridging, and regrowth moves; the latter two being connectivity altering. The regrowth move consists of regrowing up to 3 beads at either end of the chain, using a configuration-bias algorithm, and includes the possibility of reversing the chain. This MC move set was used with the WL algorithm as in [2] to iteratively approximate the density-of-states function $g(E)$, giving a well-sampled set of configurations across the whole energy range. In CD, free flight of the spheres occurs between elastic collisions in standard fashion [8, 9]. Collisions occur at each discontinuity in eqns (1), (2). Additionally, thermal jolts reselect the velocities of individual atoms from the Boltzmann distribution and introduce a stochasticity into the dynamics. The time separation of the jolts has a Poisson distribution with mean time τ giving the strength of the coupling (typically $\tau = 0.1$ in reduced units). The WL simulations were used to determine the thermodynamic freezing temperatures, T_f , at which doubly-peaked canonical ensemble energy distributions $P_C(E)$ were observed, for a range of χ values. A typical example is shown in Figure 1. Similar probability distributions were determined from the CD simulations. These confirmed a unique globule phase; we denote this as B and identify a single energy distribution $P_B(E)$. However, CD simulations initialized in different realisations of the crystal phase did not explore as wide an energy range as the corresponding WL simulations. From this we infer that the crystal phase (at T_f) consists of a large number of

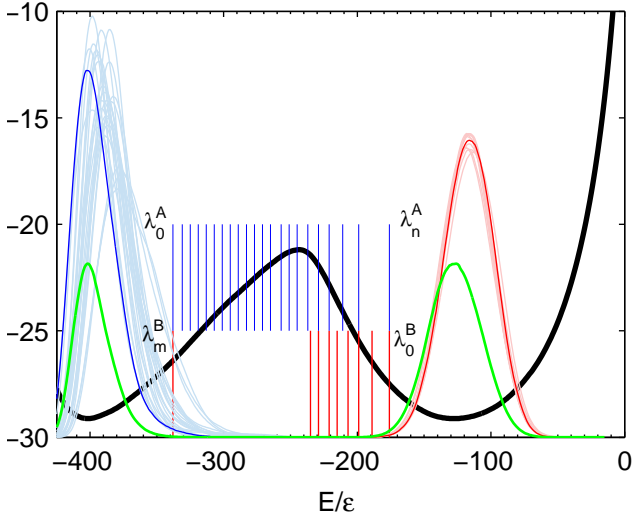


FIG. 1: Free energy $F_C(E) = E - T \ln(g(E))$ (black) and canonical probability function $P_C(E) \propto g(E) \exp(-E/k_B T)$ (green) obtained from WL simulation. Vertical lines schematize the interfaces λ_i^B in folding (red) and the interfaces λ_i^A in unfolding (blue) directions. Displayed also are energy distributions of states in the crystalline phase (blue) and in the globule phase (red); these graphs must be scaled according to the folding and unfolding rates [10] to be comparable with $P_C(E)$. The vertical axis of the probability function is not shown; these functions are normalized. The data are for the chain with $\chi = 1.07$, $T = 0.498$.

basins with slightly different mean energies, separated by kinetic barriers which cannot be overcome (at these temperatures) on a simulation timescale without the use of unphysical MC moves. We therefore identify the *crystal state A* as being the state, that can be reached from other states within the distribution function $P_A(E)$ via constant temperature MC (including unphysical moves), which has the lowest mean energy. CD alone succeeded in sampling these low-lying basins only for $\chi \geq 1.05$.

The order parameter λ describing the qualitative difference between the globule and crystal state was chosen to be equal to the potential energy E of the chain. To accelerate the sampling of fluctuations leading to the folding and unfolding transitions we used Forward Flux Sampling (FFS) [11–13], which separates the phase space by n hyperplanes orthogonal to the order parameter λ and measures the probability flux through these planes. Here, the hyperplanes (also called interfaces) and the associated energies are denoted by the same symbol λ_i . FFS for our chain was performed in both directions, i.e. from globule to crystal, and from crystal to globule. This approach allows us to focus on kinetic effects associated with each direction separately, and is better suited to our dynamics than the backwards/forwards shooting approaches of alternatives such as Transition Interface Sampling, which are otherwise essentially equivalent [14]. FFS is also known to be relatively insensitive to the

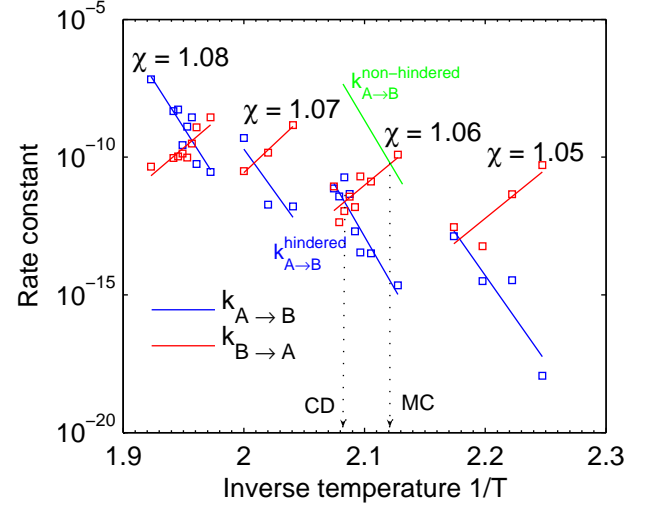


FIG. 2: Chevron plots with folding and unfolding rates computed by FFS. The intersections give the transition temperatures T_f^{CD} estimated by collision dynamics. The schematic for $\chi = 1.06$ shows that kinetic hindering of unfolding in CD simulations can explain the observation that $T_f^{\text{CD}} > T_f^{\text{MC}}$. The schematic assumes that the hindering involved in the folding process is negligible.

choice of order parameter. The rate from crystal to globule is given by:

$$k_{A \rightarrow B} = \frac{\langle \Phi_{A, \lambda_0^A} \rangle}{\langle h_A \rangle} \prod_{i=0}^{n-1} P(\lambda_{i+1}^A | \lambda_i^A). \quad (3)$$

Here $\langle \Phi_{A, \lambda_0^A} \rangle$ is the probability flux through λ_0^A ; state A is defined by an interface $\lambda_A < \lambda_0^A$. $\langle h_A \rangle$ is the probability of being most recently in state A . The fraction $\langle \Phi_{A, \lambda_0^A} \rangle / \langle h_A \rangle$ then simply represents the inverse of the average time needed to reach λ_0^A from the first crossing of λ_A . $P(\lambda_{i+1}^A | \lambda_i^A)$ is the conditional probability of reaching the interface λ_{i+1}^A from λ_i^A , which is given by the fraction of partial pathways started from λ_i^A which reach λ_{i+1}^A before they fall back to λ_A . The partial pathways are eventually connected into the full transition pathways starting in A and ending in B . The rate $k_{B \rightarrow A}$ from globule to crystal is given analogously. In this way, we gathered 8192 folding and unfolding transition pathways by CD simulation for each χ and T , choosing temperatures in the vicinity of the transition temperatures determined by WL.

The interfaces in FFS are positioned as follows. Energies λ_0^A and λ_0^B are chosen such that the ranges $\lambda < \lambda_0^A$ and $\lambda > \lambda_0^B$ capture 99.9% of the corresponding integrated densities P_A and P_B . The boundaries of A and B are then defined as $\lambda_A \equiv \lambda_0^A - 50$ and $\lambda_B \equiv \lambda_0^B + 50$. To reconstruct the full pathways we define an extra plane as $\lambda_n^A \equiv \lambda_0^B$. Interface λ_{n-1}^A is placed close to the isocom-

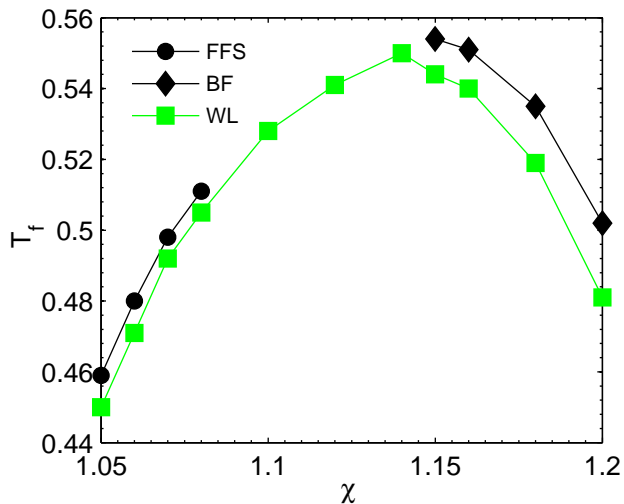


FIG. 3: Phase diagram with freezing temperatures determined from chevron plots (T_f^{CD} , FFS and BF) and by Maxwell construction (T_f^{MC} , WL). The errors are smaller than the symbol sizes.

mittor, specifically such that $P(\lambda^B | \lambda_{n-1}^A) = 0.9$. The remaining interfaces $\lambda_1^A, \dots, \lambda_{n-2}^A$ are chosen according to the optimization scheme of Ref. [13]. An analogous procedure applies to interfaces $\lambda_1^B, \dots, \lambda_m^B$.

To avoid undersampling of λ_0^B , 32 random disordered chains are equilibrated, and λ_0^B is sampled in parallel starting from these different chains. Similarly, CD simulations started from different points in λ_m^B gave $P_A^{(i)}(E)$, $i = 1, \dots, 256$. λ_0^A is then sampled in parallel starting from a configuration belonging to the distribution $P_A(E)$ having the lowest mean energy, and from 31 other configurations which are separated from this state by a large number of MC moves, including connectivity-altering moves, conducted at constant temperature.

Transition temperatures T_f^{CD} , defined dynamically by the equation $k_{A \rightarrow B} = k_{B \rightarrow A}$, are obtained from the FFS simulations via a so-called chevron plot, Fig. 2 (based on the assumption of Arrhenius-like behaviour). These were slightly, but systematically, higher than the WL transition temperatures T_f^{MC} by an amount $\Delta T = T_f^{CD} - T_f^{MC} = 0.005\text{--}0.012$ (Fig. 3). For the larger values of χ , barrier crossing could be observed directly, and rate constants were calculated by brute force (BF) simulation. Transition temperatures obtained in this way were again comparable to the temperatures obtained in MC simulation, but were also systematically higher. This suggests that the discrepancy is due to a real dynamical (kinetic) effect rather than any deficiency of the FFS algorithm itself. The most likely explanation is that CD, hindered by metastable basins acting as kinetic traps, yields significantly lower unfolding rates than the MC simulation, which escapes the traps with the aid of non-kinetic connectivity-altering moves. Fig. 2 indicates schemati-

cally the extent to which the unfolding regression line must be shifted to give the observed shift ΔT ; this could provide a quantitative measure of the kinetic hindering.

The remaining part of the paper points out a remarkable property the *dynamical matrix* of our polymer system in the vicinity of the free energy maximum. We focus on the chain with $\chi = 1.07$ and $T = 0.498$, but similar results were obtained for other parameters, becoming even more distinct with decreasing χ . The dynamical matrix of our chain is defined as follows:

$$\Gamma_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } r_{ij} \leq \chi\sigma, \\ 0 & \text{if } i \neq j \text{ and } r_{ij} > \chi\sigma, \\ -\sum_{i, i \neq j} \Gamma_{ij} & \text{if } i = j. \end{cases} \quad (4)$$

It is closely related to the contact (or adjacency) matrix, whose elements are unity for atom pairs within interaction range, and zero otherwise. Contact matrices have been used to describe the equilibrium structure of proteins in terms of amino acid contacts [15, 16]; as described by Bahar et al. [17] and Sadoc [18] this idea may be extended, through the dynamical matrix, to give a simple network model of vibrational motions. Here we suggest that the topology of the interactions in the chain around the critical point of collapse, described by Γ , may be linked to the dynamical effects that we have observed.

The largest eigenvalue of Γ will be denoted as γ . Let $a(\lambda; \gamma)$ be the probability distribution of γ at an interface λ sampled by pathways started from A , and $b(\lambda; \gamma)$ the same quantity but sampled by pathways started from B . We found that $a(\lambda; \gamma)$ is unimodal (approximately Gaussian) at all interfaces, with the mean value growing with λ , and that the conformations at λ with large γ are more likely to crystallise. The distribution $b(\lambda; \gamma)$ at interfaces far enough from the isocommittor is also unimodal with similar properties, but becomes bimodal at the interfaces λ close to the isocommittor. The critical value separating these two modes is denoted as γ^c . The insets of Fig. 4 show that pathways started from A , and reaching these energies, do not sample the population of low-eigenvalue states. The microscopic reversibility of our dynamics then implies that folding transition pathways must cross λ_{n-1}^B at $\gamma > \gamma^c$. Indeed, the probability analysis in Fig. 4 shows that pathways started at λ_{n-1}^B with $\gamma < \gamma^c$ have almost no chance to reach A . The difference between the distributions $a(\lambda; \gamma)$ and $b(\lambda; \gamma)$ confirms our suspicion that the forward-going and reverse-going ensembles of reactive trajectories are not identical (due to kinetic hindering), and that the top eigenvalue of the dynamical or contact matrix may be able to discriminate between these families of trajectories. Incidentally, the *equilibrium* distribution of γ at this energy, obtained by WL, is very similar to the nonreactive $\gamma < \gamma^c$ portion of Fig. 4(b). Similar observations apply to other eigenvalues near the top of the spectrum of Γ . It is significant that typical $\gamma > \gamma^c$ configurations appear to have a com-

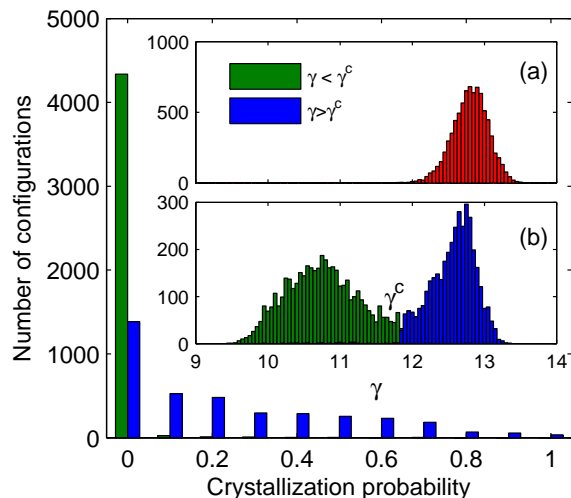


FIG. 4: Configurations at λ_{n-1}^B with largest eigenvalue γ of the dynamical matrix lower than the critical value ($\gamma^c = 11.9$) have almost no chance to crystallize. Insets: (a) Unimodal distribution $a(\lambda^A; \gamma)$ at the surface $\lambda^A = -235$ sampled by pathways started in A. (b) Bimodal distribution of $b(\lambda^B; \gamma)$ at the same surface ($\lambda_{n-1}^B = -235$) but sampled by pathways started in B.

pact crystal nucleus with attached chains or loops, while for $\gamma < \gamma^c$ the same number of interactions are typically arranged in a single, less well ordered, cluster.

It is worth mentioning that two structures on surfaces λ close to the isocommittor have also been identified by Taylor et al. [5] using the radius of gyration (R_g) as a second reaction coordinate. This result was also confirmed here. The correlation between R_g and γ was tested and found to be only weak. An analysis similar to that in Fig. 4 showed that γ has significantly better predictive properties than R_g .

Why do we believe that the dynamical matrix, and the associated contact matrix, deserve further study? As mentioned above, they give a rather general connection between the topology defined by the interactions within a chain configuration and its dynamical evolution, in the approximation of an elastic network model. This has not only been used in the discussion of proteins to identify vibrational modes of oscillations [17, 18], but also in the definition of nodes and an order parameter (a distance between nodes) in dynamical network models of the folding process itself [19]. The dynamical matrix is also being used for the description of glassy structures in colloidal systems [20]. Most recently, the contact matrix of an atomic cluster has been used as a generator of order parameters for metadynamics simulations [21]. Our results clearly reinforce the view that the dynamical matrix is a simple object capturing successfully important topological or vibrational features of various interacting systems.

To summarize, the transition of the homopolymer chain from the disordered globule to the crystal state

has been simulated by dynamical forward flux sampling and brute force simulation. The results gave strong evidence that kinetic effects play an important role in the determination of the effective transition temperature. We then showed that the eigenvalues of the dynamical matrix yield further important information regarding the forward and reverse trajectories in the folding transition which complement the potential energy as an order parameter.

Computer facilities were provided by the Centre for Scientific Computing at the University of Warwick. We thank Dr. Rosalind Allen, Prof. Kurt Binder, Dr. Ellak Somfai, and Adam Swetnam for useful discussions.

-
- [1] J. E. Magee, V. R. Vazquez, and L. Lue, Phys. Rev. Lett. **96**, 207802 (2006).
 - [2] M. P. Taylor, W. Paul, and K. Binder, Phys. Rev. E **79**, 050801 (2009).
 - [3] M. P. Taylor, W. Paul, and K. Binder, J. Chem. Phys. **131**, 114907 (2009).
 - [4] D. T. Seaton, T. Wüst, and D. P. Landau, Phys. Rev. E **81**, 011802 (2010).
 - [5] M. P. Taylor, W. Paul, and K. Binder, Phys. Procedia **4**, 151 (2010).
 - [6] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications* (Academic Pr, 2002).
 - [7] D. P. Landau and K. Binder, *Guide To Monte Carlo Simulations In Statistical Physics* (Cambridge Univ Pr, 2009).
 - [8] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids* (Oxford University Press, 1987).
 - [9] D. C. Rapaport, *The art of molecular dynamics simulation* (Cambridge Univ Pr, 2004).
 - [10] C. Valeriani, R. J. Allen, M. J. Morelli, D. Frenkel, and P. R. ten Wolde, J. Chem. Phys. **127**, 114109 (2007).
 - [11] R. J. Allen, P. B. Warren, and P. R. ten Wolde, Phys. Rev. Lett. **94**, 018104 (2005).
 - [12] R. J. Allen, C. Valeriani, and P. R. ten Wolde, J. Phys. Condens. Matter **21**, 463102 (2009).
 - [13] F. A. Escobedo, E. E. Borrero, and J. C. Araque, J. Phys. Condens. Matter **21**, 333101 (2009).
 - [14] P. G. Bolhuis, J. Chem. Phys. **129**, 114108 (2008).
 - [15] M. Vendruscolo, B. Subramanian, I. Kanter, E. Domany, and J. Lebowitz, Phys. Rev. E **59**, 977 (1999).
 - [16] J. L. England and E. I. Shakhnovich, Phys. Rev. Lett. **90**, 218101 (2003).
 - [17] I. Bahar, A. R. Atilgan, M. C. Demirel, and B. Erman, Phys. Rev. Lett. **80**, 2733 (1998).
 - [18] J. F. Sadoc, Eur. Phys. J. E **18**, 321 (2005).
 - [19] G. Lois, J. Blawdziewicz, and C. S. O'Hern, Phys. Rev. E **81**, 051907 (2010).
 - [20] K. Chen, M. L. Manning, P. J. Yunker, W. G. Ellenbroek, Z. Zhang, A. J. Liu, and A. G. Yodh, arXiv:1103.2352v1 (2011).
 - [21] F. Pietrucci and W. Andreoni, Phys. Rev. Lett. **107**, 085504 (2011).